DOCUMENT RESUME

ED 180 195 FL 010 677

AUTHOR Backman, Jarl
TITLE Subjective Structures in Linguistic Recurrence. Educational Reports Umea, No. 19.
INSTITUTION Umea Univ. (Sweden).
SPONS AGENCY Swedish Council for Social Science Research, Stockholm.
PUB DATE 78
NOTE 22p.
AVAILABLE FROM Department of Education, University of Umea, S-901 87 Umea, Sweden (free)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Processes: Componential Analysis: *Comprehension: Grammar: Language Research: Lexicology: Linguistic Theory: *Phrase Structure: *Psycholinguistics: *Psychology: Swedish: Vocabulary: *Word Frequency

ABSTRACT

Three experiments were conducted to test the psychological relevance of objectively quantified word collocations. The first experiment showed that perceived frequency of word combinations roughly followed the objective count. Another recurrent quality of words, constructional tendency, was supplemented as independent variable in the two following experiments. This variable reflects a word's tendency to appear in word combinations and it was found to interact with frequency when subjects rated frequency and comprehensibility. The experiments showed that word collocations, defined at the levels of combinations and constructional tendency of individual words, can be supposed to have psychological counterparts. The linguistic recurrence seems to have cognitive representations. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED180195

educational reports umeå

no 19 1978

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Umea Univ.
Dept. of Education

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FL010677

Backman, J. Subjective Structures in Linguistic Recurrence. Educational Reports Umeå, 1978, No 19 ×

The tendency of words to occur repetitively in combinations with other words is called linguistic recurrence. Examples of such word combinations are *"to all appearances"*, *"as to the rest"* or *"may have"*. Studies of such word collocations belong to phraseology which has been described as an intersection between lexicon and grammar (Allén, 1973). This field is, however, extremely relevant for psycholinguistics as well. The predominant atomistic single-word theories present no attempts at dealing with recurrence. The prediction from these theories have to be based on taking the units in the phrases separately. It seems, however, reasonable to assume that phrases have specific semantic representations of their own which cannot be directly derived from the included words. The phrase *"to all appearances"* will probably not have three separate entries in the semantic memory of individuals. Such a representation would seem uneconomical and almost inconceivable in language production. Miller's familiar quotation *"the meaning of an utterance is not a linear sum of the meanings of the words that comprise it"*, (Miller, 1965, p. 18) can probably be applied to a phrase as well. Allén (1973) also seems to advocate some kind of block representation. This is probably particularly obvious for a special type of phrases namely idioms. These expressions have meanings which are not predictable from their individual elements, e.g. *"to be on the carpet"*.

Allén (1975) has, in a recent work, which is unique in the world, made a list of word combinations occurring

× This research has been supported by The Swedish Council for Social Research.

in Swedish. The list is made on three main levels of "description":

1) combinations for which the criterion is mainly statistical. All word combinations having at least two identical entries in a million words have been listed.
2) constructions, where the criteria are linguistically relevant. It must be possible to classify the word collocations in the grammatical categories noun phrases, verb phrases, connectives, and clauses.
3) idioms, where the combinative meaning is not predictable from the individual words.

The purpose of this work is to test the psychological relevance of serial word occurrences. It has earlier been shown that words have several subjective attributes and that some of them seem to be directly related to objective frequency. It was for example found that subjective frequency had a very high correlation (r=.92) with earlier lists made by Allén (1970; 1971) (Backman, 1976). Such a high degree of correlation gives, in addition to the above-mentioned reasons for restrictions on semantic representations, cause to assume a psychological counterpart to objectively quantified recurrence. The experiments presented below are of an explorative nature since there is no prior empiric on linguistic recurrence and its psychological counterpart.

EXPERIMENT I

Verbal Material

The aim of the first experiment was to test two variables on Allén's (1975) first level of description, the level of word combinations. The first variable is the frequency of the combination (OF), the second is the frequency of the combination divided by a theoretically expected random-generated frequency (OF/T). These values are determined for combinations where F > 10. The theoretical value T is obtained via the formula

$$T = \frac{OF_1 \cdot OF_2 \cdots OF_n}{N^{n-1}}$$

where OF_1 is the frequency of the first word in the combination, OF_2 that of the second word, etc. N indicates the number of running words in the material and n number of words in the combination. The calculation of the quotient OF/T results in a measure of the "usualness" of the occurrence in relation to the random occurrence for a combination. The range is found to lie between the extreme values $1.8 \cdot 10^{8}$ (*"in a wholly different way"*) and $3.1 \cdot 10^{-2}$ (*"one that; one to"*). The first value represents the most frequent and the second the least frequent combination. Arbitrary equidistant intervals were fixed as there are no earlier studies with a psychological approach in "quantitative phraseology" and consequently no subjective metrics in the above-mentioned respects. The levels ≈ 10, ≈ 60, and ≈ 110 were chosen for OF and for the OF/T-variable the exponents approximatively 1, 4, and 7. The

units *"and so on"*, *"that is"*, and *"up to and including"* on the exponent level 7 are not as in Allén represented by their abbreviations (the corresponding expressions in Swedish are *"och så vidare"*, *"det vill säga"*, and *"till och med"* abbreviated *"osv."*, *"dvs."* and *"tom."*). OF has here, however, been calculated as the sum of the abbreviation entries.

Procedure

The variables give with the three levels a 3 x 3 design and each combination of levels was represented by two words. The experiment thus included 18 different word combinations, presented in Table I. The number of words in the combinations was three (in Swedish) in all cases except in one at OF ≈ 60 and OF/T - exponent 7.

All word combinations were presented in randomized order and the subjects were asked to estimate the frequency of each combination in relation to a standard. This standard was given the numerical value 100 and consisted of the combination *"AT FIRST HAND"*. The technique was thus Steven's method of magnitude estimation which has also been used for scaling the subjective frequency of single lexical items (Backman, 1976; Carroll, 1971).

Subjects

The experimental group consisted of fifteen students of pedagogics (B1-level) at the University of Umeå who had not before taken part in similar experiments. The test was carried out during ordinary lesson time.

Table I. Word combinations on different levels of frequency and OF/T-exponents

Frequency	OF/T-exponent 10^1	10^4	10^7
≈ 10	*om man så/ if you so*	*av olika slag/ of various kinds*	*under årens lopp/ in the course of time*
	kanske är det/ it may be	*grund och botten/ at heart*	*att ägna sig/ to devote oneself to*
≈ 60	*det finns en/ there is one*	*i vårt land/ in our country*	*och så vidare/ and so on*
	det är också/ it is also	*en hel del/ a great deal*	*av allt att döma/ to all appearances*
≈ 110	*det är ett/ it is a*	*i själva verket/ in reality*	*det vill säga/ that is*
	är det inte/ is it not	*i stort sätt/ on the whole*	*till och med/ up to and including*

RESULTS

Data are presented graphically in Figure I which illustrates a comparatively high degree of correspondence between frequency levels over OF/T-exponents. The frequency levels group themselves ordinally so that the highest level (≈ 110) is also perceived as the most frequent one.

An ANOVA was made in order to get more information on the relations between the recurrence variable. This resulted

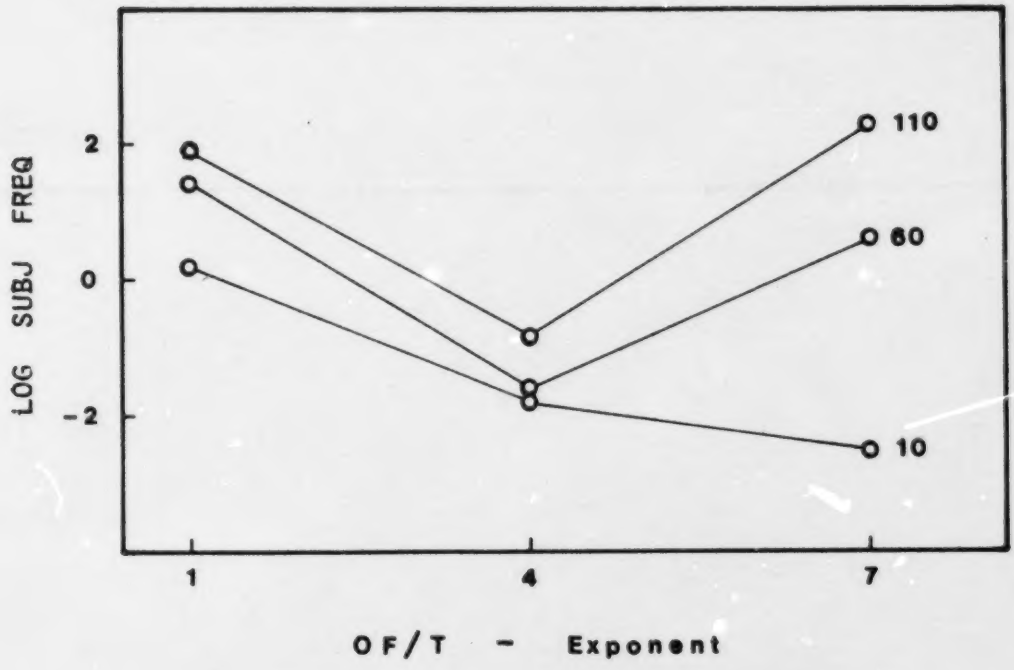

Figure I. Logarithmic subjective frequency values for different frequencies (10, 60 and 110) at different OF/T-exponents.

in an significant interaction $\underline{F}$ (4,9) = 1.24 $\underline{MS}_e$ = 0.44, $\underline{p}$ > .10. The two main effects were found to be significant. The calculation for $\underline{OF}$ yielded $\underline{F}$ (2,9) = 5.36, $\underline{MS}_e$ = 0.44, $\underline{p}$ < .05 and for $\underline{OF/T}$ the corresponding figures were $\underline{F}$ (2,9) = 5.72, $\underline{MS}_e$ = 0.44, $\underline{p}$ < .025.

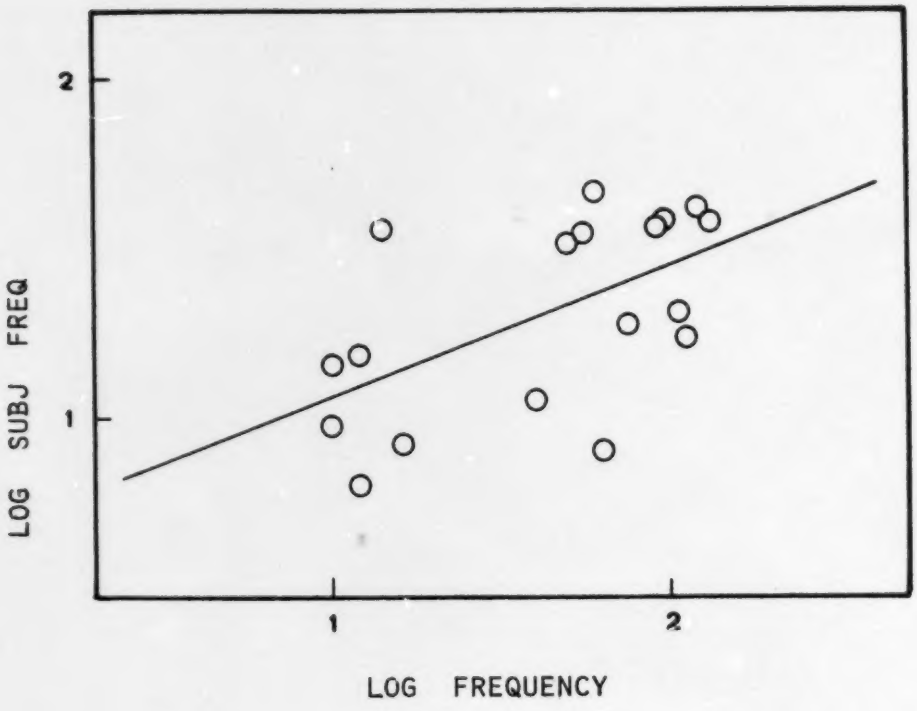

Figure 2. Logarithmic subjectiv frequency plotted against logarithmic objective frequency. The straight line is fitted by means of the least square criterion.

Figure 2 present logarithmic subjective frequency estimates in relation to objective frequency. A straight line has been fitted to the data and the relation can be described by a power function the exponent of which was calculated to 0.39. The correlation expressed in a product moment coefficient was 0.56.

EXPERIMENT II

The following two experiments included the second of Allén's (1975) levels of description. A number of linguistic selection criteria have been applied to the first level (combinations) and they have yielded the subset of constructions. On this level Allén introduces a variable called constructional tendency (Kb), which can be expressed in a proportion: the frequency of constructions of a word / the total (graphic word) frequency of the word shown as percentages in Allén). The recurrent quality of a word is thus reflected in the Kb-measure. The closer Kb comes to 1.0, the more often it tends to appear in word collocations. A variation in the variables Kb and OF was chosen for Experiment two. Four levels were chosen on the former variable: 0.20, 0.40, 0.60, and 0.80. The OF-variable was also represented on four levels, namely 10, 50, 90 and 130. The selection of words was not restricted regarding for exampele syntactic category of construction criteria. The word length in number of letters was, however, to be approximately the same. The words included in Experiment II are presented in Table II.

The purpose of this second experiment on recurrence is to find out if the constructional tendency of words has any psychological relevance. This is done by examining whether estimation of comprehension is related to Kb.

Table II. Specification of costructional tendency (Kb) and objective frequency (OF) for the words in experiment II.

	OF			
Kb	10	50	90	130
.20	*STREJKEN/ THE STRIKE*	*SÖKTE/ SEARCHED*	*SPECIELLA/ SPECIAL*	*KOSTAR/ COSTS*
.40	*BEVARAT/ PRESERVED*	*HEKTAR/ HECTARE*	*VISSTE/ KNEW*	*RELATIVT/ RELATIVELY*
.60	*KONSUL/ CONSUL*	*AKTUELLT/ ACTUAL*	*BIDRAG/ CONTRIBUTION*	*OMRÅDET/ THE AREA*
.80	*UPPEHÖLL/ MAINTAINED*	*RISKEN/ THE RISK*	*HÖSTENS/ OF THE AUTUMN*	*SKRIVA/ WRITE*

Procedure

One word was chosen on each Kb- and OF-level and the experiment thus included 16 (4 x 4) words. These were introduced to the subjects in a randomized order. The task of the subjects was, for the first group, to underline that word which was considered to occur most often and, for the second group, to underline that word which was easiest to understand. The technique was thus Thurstone's paired comparisons.

Subjects

The subjects were 52 pupils from upper secondary school. Half of them estimated subjective frequency (SF), and

the other half estimated comprehension (C). The experiment was carried out during ordinary lesson time.

RESULTS

Calculated proportions were transformed into angular deviates and an ANOVA yielded a significant interaction for SF between OF and Kb: F (9,240) = 10.16 MS_e = 0.38, $p < .001$. The main effect of OF resulted in F (3,240) = 14.18, $p < .001$ and for Kb in F (3,240) = 9.08, $p < .001$. The main effect of OF resulted in F (3,240) = 14.80, $p < .001$ and for Kb in F (3,240) = 9.08, $p < .001$. The analysis of C also resulted in a significant interaction between OF and Kb: F (9,240) = 13.42, MS_e = 0.26, $p < .001$ and for Kb, F (3,240) = 13.49, $p < .001$. The results for the two dependent variables were thus identical as regards significance for main effects and interactions. The judgment dimensions are, however, not identical which is clear from the calculated correlation coefficient: r = .64. The upper figure of figure 3 shows single factorially how SF grows clearly linearly with OF, while C shows a tendency to an asymptot at the highest frequencies. Both variables yield a U-function towards Kb, which is clear from the lower figure. The two functions are here strikingly alike: only the min. - points differ, but only slightly. The interactive conditions should, however, be kept in mind when reading the figures.

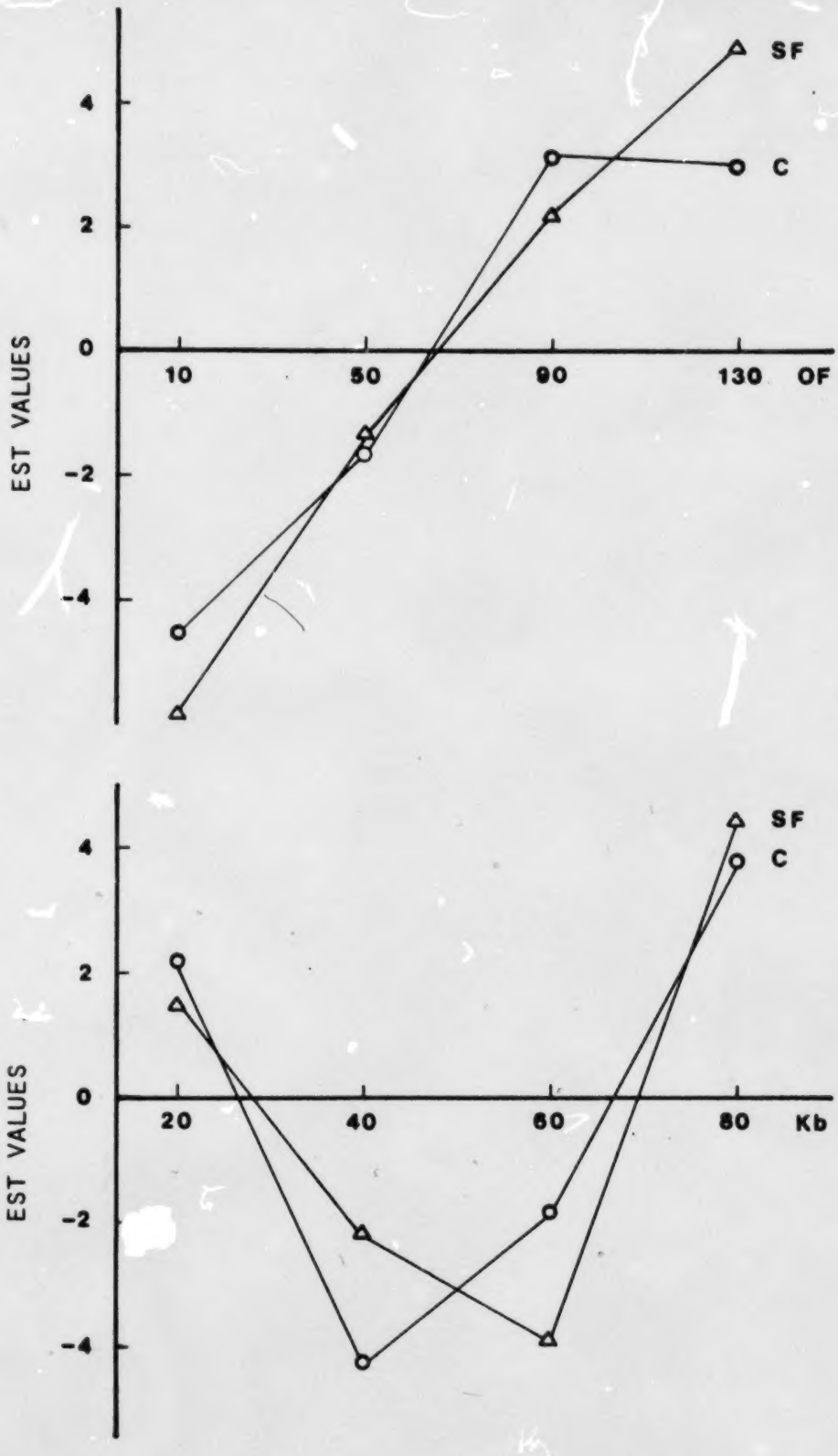

Figure 3. Subjective frequency (SF) and comprehension (C) as function of (column sums in the proportion matrices) objective frequency (OF, the upper diagram) and constructional tendency (Kb, the lower diagram).

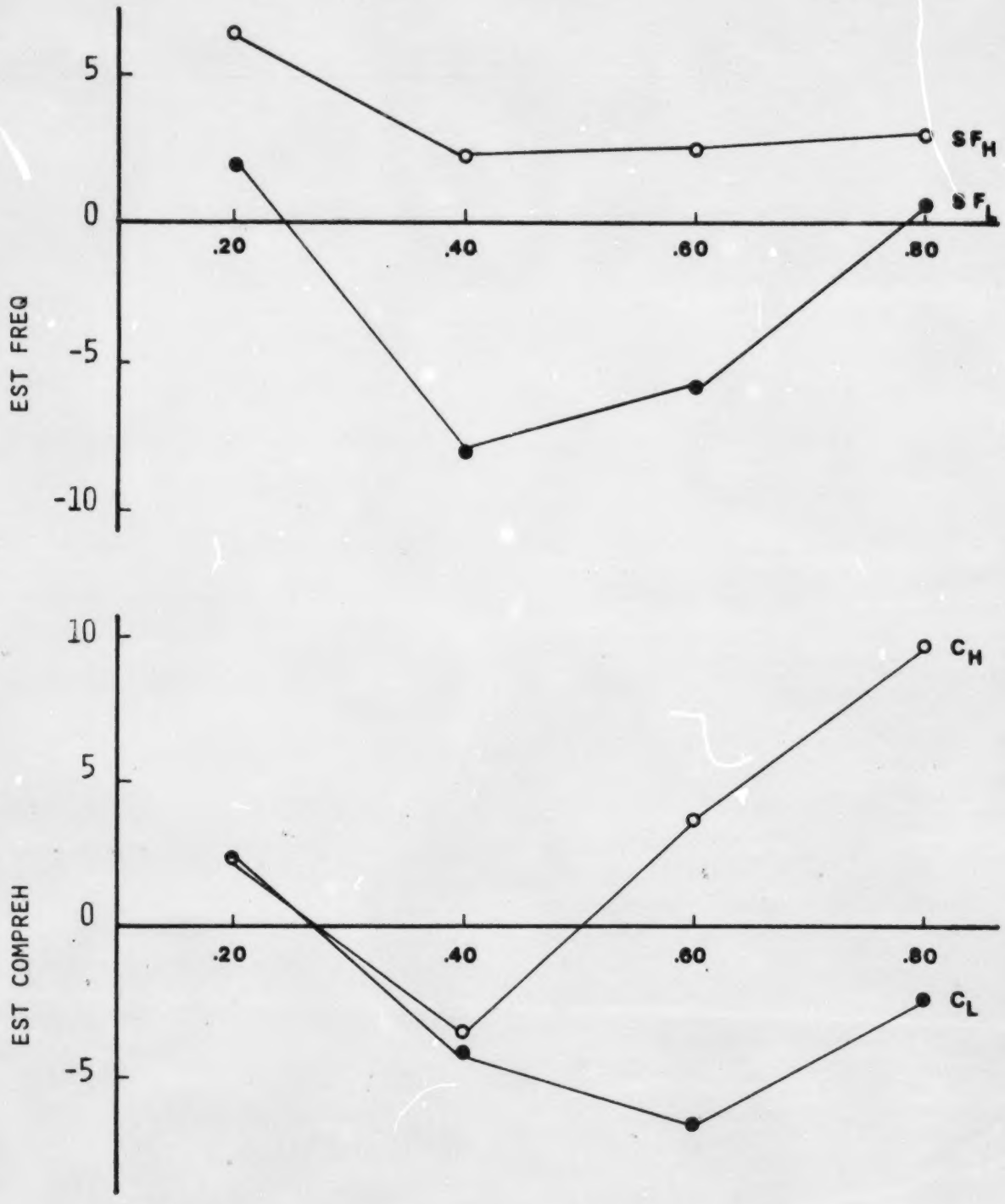

Figure 4. Subjective frequency for the highest frequency level (SF_H) and the lowest (SF_L) on different levels of constructional tendency (the upper diagram). The lower diagram shows the corresponding for rated comprehension (C).

Figure 4 illustrates the two dependent variables in relation to Kb. The two lowest and the two highest objective frequencies have here been put together into a low and a high frequency level (index L and H respectively). The upper figure shows that the low- and high frequency levels are perceived as clearly distinctive, whereas the U-relation is most strongly marked for the lowest frequency level. No difference is noted between the low- and high frequency levels at the two lowest Kb-values for the comprehension dimension. The two highest, on the other hand, show a great difference between the frequency levels.

EXPERIMENT III

The third experiment was similar to the previous one, but with a change in the OF- and BV-variable. The aim was here to investigate a larger range regarding frequency. Three levels were chosen, equidistantly separated through a logarithmic unit. Since there were still no restrictions on the selection of words, the words on the highest frequency level were exclusively made up of function words. The estimated words are presented in Table III.

The more interesting comprehension dimension was chosen of the two dependent variables used the previous experiment. The purpose was to examine whether the relations between OF and Kb found in Experiment II would remain in an extension of the objective frequency range. The method and procedure were identical with those used in Experiment II. The subjects were 32 students at Umeå University.

Table III. Verbal material in Experiment III specified in objective frequency (OF) and constructional tendency (Kb).

	OF		
Kb	30	300	3000
.20	*VÄNTADE/ WAITED*	*HENNES/ HER*	*ELLER/ OR*
	HUSEN/ THE HOUSES	*SADE/ SAID*	*NU/ NOW*
.40	*VÄRME/ VARMTH*	*LITEN/ SMALL*	*VID/ AT*
	TONEN/ THE TONE	*FÖRE/ BEFORE*	*UNDER/ UNDER*
.60	*SAND/ SAND*	*OMKRING/ AROUND*	*SÅ/ SO*
	TILLVARON/ THE EXISTENCE	*BORT/ AWAY*	*MOT/ TOWARDS*
.80	*LÄNGDEN/ THE LENGTH*	*PLATS/ PLACE*	*VAR/ WAS*
	FLYGA/ FLY	*LANDET/ THE COUNTRY*	*ÅR/ YEAR*

RESULTS

The outcome of an ANOVA yielded the same significanses as before. For the interaction OF x Kb we obtained F (6,276) = 12.50, MS_e = 0.08, $p < .001$, for the main effect of objective frequency F (2,276) = 172,5, $p < .001$ and for constructional tendency F (3,276) = 59.26, $p < .001$.

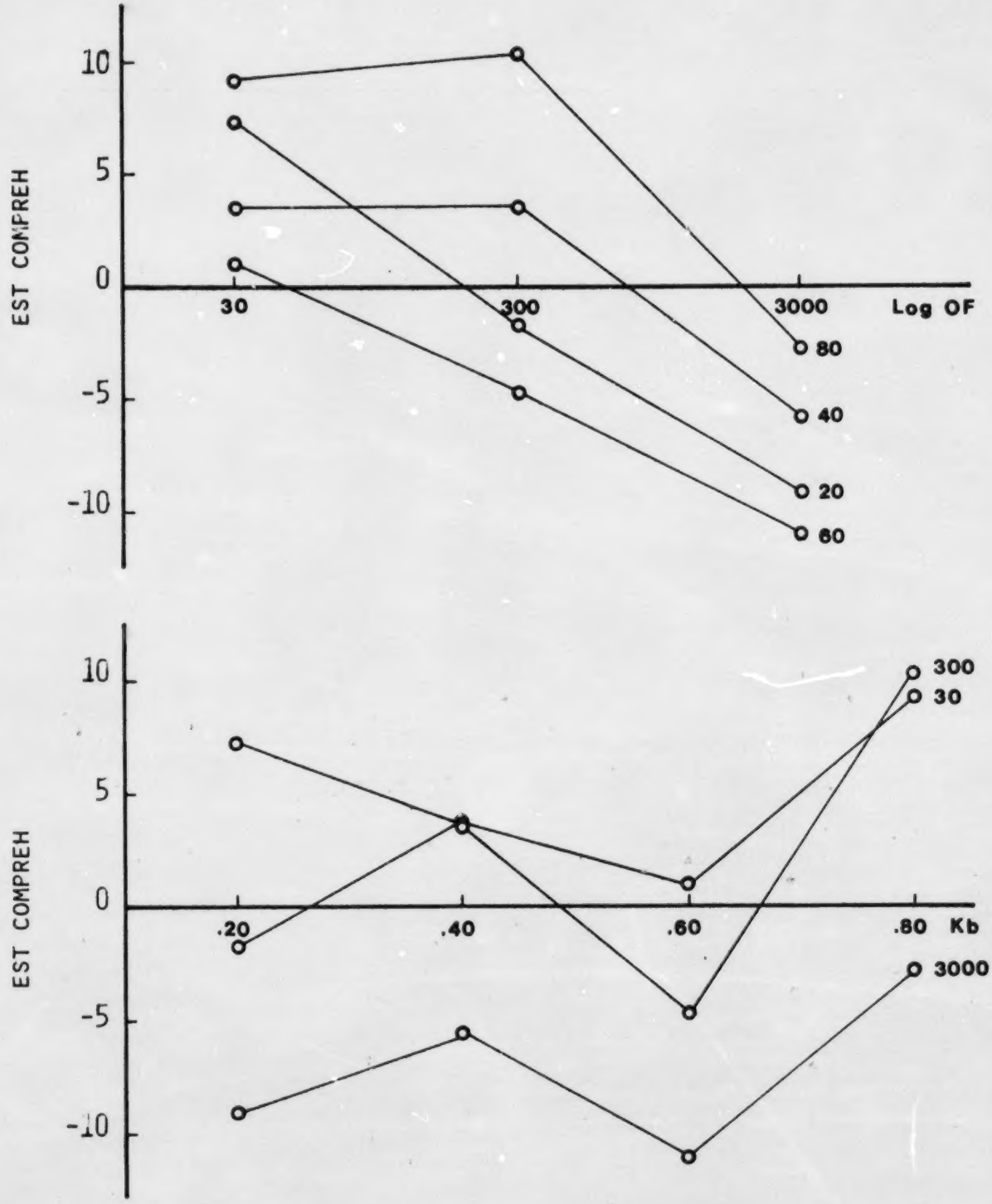

Figure 5. Rated comprehension (column sums) on different frequency levels with constructional tendency as parameter (the upper diagram). The lower diagram shows the same but with frequency as parameter.

For reasons mentioned before, there was no further ANOVA-testing. Tendencies can, however, be seen in figure 5. The upper figure shows that comprehension seems to be lowest for the highest frequency level. The lower figure shows that the highest Kb-level yields the highest comprehension on every frequency level. The earlier indicated U-function can now only be discerned on the lowest frequency level.

SUMMARY AND DISCUSSION

The first experiment on linguistic recurrence, made on the level of combinations, showed that individuals have a subjective representation of phrases which roughly follows an objective quantification. Subjective frequency of word combinations is, as for single-word exposures, related to objective frequency like on the whole to the measure of "usualness" (OF/T) introduced by Allén (1975). It is also interesting to observe that when subjective frequency is related to objective frequency the relation can be described as a power function with the exponent n = 0.39. The agreement with single-word estimations is noteworthy. The exponent n = 0.35 has been presented in an earlier report (Backman, 1976).

The variable constructional tendency was tested in the second experiment and it was found to interact with frequency when individuals rated subjective frequency as well as comprehension. Both of the dependent variables tended to be parabolically related to constructional tendency on the lowest frequency levels. This could also be observed in the third experiment where the range in objective fre-

quency had been extended to comprehension measurements.

The above experiments should be regarded as explorative since there were no restrictions on the selection of verbal material other than for the selection of levels of independence. The experiments show, however, that word collocations, phrases, and the constructional tendency of individual words can be supposed to have psychological counterparts, that linguistic recurrence seems to have cognitive representations. Continued research should aim at systematizing the selection of verbal material regarding for example the selection criteria for constructions in order to confirm the U-function which was found between comprehension and constructional tendency.

REFERENCES

Allén, S. Frequency Dictionary of Present-Day Swedish Based on Newspapaer Material. 1. Graphic Words, Homograph Components. Stockholm: Almqvist & Wiksell, 1970.

Allén, S. Frequency Dictionary of Present-Day Swedish Based on Newspaper Material. 2. Lemmas. Stockholm: Almqvist & Wiksell, 1975.

Allén, S. Frequency Dictionary of Present-Day Swedish Based on Newspaper Material. 3. Collocations Stockholm: Almqvist & Wiksell, 1975.

Allén, S. On phraseology in lexicology. Cahiers de Lexicologie, 1976, 29, 83-90.

Backman, J. Some common word attributes and their relations to objective frequency counts. Scandinavian Journal of Educational Research, 1976, 20, 175-186.

Carroll, J.B. Measurement properties of subjective magnitude estimates of word frequency. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 722-729.

Miller, G.A. Some preliminaries to psycholinguistics. American Psychologist, 1965, 20, 722-729.

EDUCATIONAL REPORTS, UMEÅ UNIVERSITY AND UMEÅ SCHOOL OF EDUCATION

Earlier reports in this series:

1973

1. READABILITY OF STAVES WITH DIFFERENT LINE AND NOTE DISTANCES.
Jarl Backman and Allan Lindahl

2. LITERACY AND SOCIETY IN A HISTORICAL PERSPECTIVE - A CONFERENCE REPORT.
Egil Johansson (Ed.)

3. THEORETICAL PROBLEMS IN CONSTRUCTION OF CRITERION-REFERENCED TESTS.
Ingemar Wedman

4. RELIABILITY, VALIDITY, AND DISCRIMINATION MEASURES FOR CRITERION-REFERENCED TESTS.
Ingemar Wedman

1974

5. DROPOUT AND PASS-RATE IN HIGHER EDUCATION - ARE THEY USEFUL MEASURES OF EFFICIENCY?
Inga Elgqvist-Saltzman

6. SOME COMMON WORD ATTRIBUTES AND THEIR RELATIONS TO OBJECTIVE FREQUENCY COUNTS.
Jarl Backman

7. ON TEST-WISENESS AND SOME RELATED CONSTRUCTS.
Ingvar Nilsson and Ingemar Wedman

1975

8. THE OCCURRENCE OF TEST-WISENESS AND THE POSSIBILITY OF INDUCING IT VIA INSTRUCTION - A cross-sectional study.
Ingvar Nilsson

1976

9. THE FUNCTION OF GROUP-SIZE AND ABILITY LEVEL ON SOLVING A MULTIDIMENSIONAL COMPLEMENTARY TASK.
Thor Egerbladh

10. READING COMPREHENSION AND PERCEIVED COMPREHENSIBILITY OF LEXICAL DENSITY AT DISCOURSE AND SENTENCE LEVEL.
Jarl Backman

11. EFFECTS OF GROUP LEARNING IN GRADE FOUR.
Thor Egerbladh

1977

12. THE HISTORY OF LITERACY IN SWEDEN. In comparison with some other countries.
Egil Johansson

13. LITERACY SCHOOLS IN A RURAL SOCIETY. A Study of Yemissrach Dimts Literacy Campaign in Ethiopia.
Margareta Sjöström and Rolf Sjöström

1977

14. TASK DIFFICULTY ABILITY LEVEL AND GROUP SIZE: A TEST OF THE THEORY OF SOCIAL DECISION SCHEMES.
Thor Egerbladh.

1978

15. WORD COMPREHENSION AS FUNCTION OF SYNTACTIC CATEGORY OF CONTENT WORDS
Jarl Backman

16. PERCEPTION OF POLYSEMY
Jarl Backman

17. SEMANTIC VARIABILITY AND WORD COMPREHENSION
Jarl Backman

18. SENTENCE PRODUCTION WITH HOMONYMOUS AND POLYSEMOUS WORDS
Jarl Backman